On the Sustainability of Web Systems Evolution

Andres Baravalle, University of East London

Cornelia Boldyreff, University of Greenwich, UK

Andrea Capiluppi, Renata Marques, Brunel University, UK)

Outline

- Introduction
- Products
- Processes
- People

Introduction

- The <u>study of web systems</u> and their evolution can be seen has having grown along <u>three main phases</u>:
  - Technologies and infrastructure
  - Web pages and client-side applications
  - Collaboration and users as content creators

Technologies and infrastructure

- Web servers, application servers, reverse proxies etc
  - Their performance
  - Capacity planning
  - How to optimise workload

Web pages and client-side applications

- The evolution of web pages
- How to develop sites by reusing code and content
- How to re-engineer the structure and links of the pages
- Best practice in the development of new content
- Testing – what to test and when

Collaboration and users as content creators

- Users as content creators (Web 2.0)
- How the developer and user communities, the stakeholders and the whole web system ecosystem have an influence on the axes of <u>products</u>, <u>processes</u> and <u>people</u>.

Products

- While Netscape and Microsoft and later on the Mozilla Foundation and Microsoft were fighting for browser predominance, technologies arose to change the face of the web to shape it as we know it now

Code reuse and abstraction

- A small number of very popular libraries gained prominence in those years – e.g. FormMail (1995) and phplib (1998)

- The issues related to maintainability were also reflected in:
  - The rise of software as WML (Website Meta Language, created in 1996)
  - The use of Server Side Includes (included in Apache in 1995)
  - The use of content management systems (Slashdot was created in 1997)

Files and URLs part ways

- Web technologies as mod rewrite (circa 1989) and session-based URLs were also popularised in the 1990s, and allowed to move away from the static relationship between files and URLs

The rise of web applications

- Geocities (1995) and Hotmail (1996)
- phpNuke and SquirrellMail (1999)
- Companies of any size started offering complex services

Popularity and reachability

- RSS (1995) allowed blogs and web sites in general to increase their popularity through syndication

- LinkExchange and LinkTrade were founded in 1996 and 1997; at his peak link exchange reached more than half of Internet-enabled households.

Open Issues

- Security: the bio-diversity that characterised the early days of the web is disappearing. Most web sites are built on top of the same set of technologies and using similar software stacks

- Niche technologies keep mushrooming – e.g. "cloud hosting", "virtual servers", "no-sql databases" and "Content Delivery Networks"

- Well be probably be seeing further separation between the large majority of the web sites using the same software stack and innovators testing and popularising new paradigms

Processes

- Traditional software development processes and methods needed rethinking
  - Metrics used to access web products and processes needed some adaptation to be applicable; new metrics such as "depth of linking" were devised
- Need for systematic development and maintenance processes for web systems

Influences

- Hypermedia design methods being promoted were derived largely from Object Oriented and database design approaches

Web metrics

- Working closely with small and medium developers allowed to:
  - Evaluate web application quality
  - To describe change
  - To guide our research on process models for web engineering
- Web site engineering stands now as a new discipline within Software Engineering

Open issues

- The heterogeneity of system components, methods and processes has led to a multi-dimensional evolution

- Web system often integrate different technologies, multiple components and web services developed and maintained by diverse teams, each component and service manifesting its own evolutionary path
  - This presents a multitude of paths to understand and attempt to manage

People

- In the last years, user-generated content has become substantial in quantity, relevant in quality, and centered around major topics and websites:
  - Multimedia (e.g. YouTube, Flickr)
  - Expert knowledge on specific topics (e.g. StackOverflow, IMDB, Wikipedia)
  - Source code released with open licenses through well known OSS repositories (SourceForge, Google Code etc).
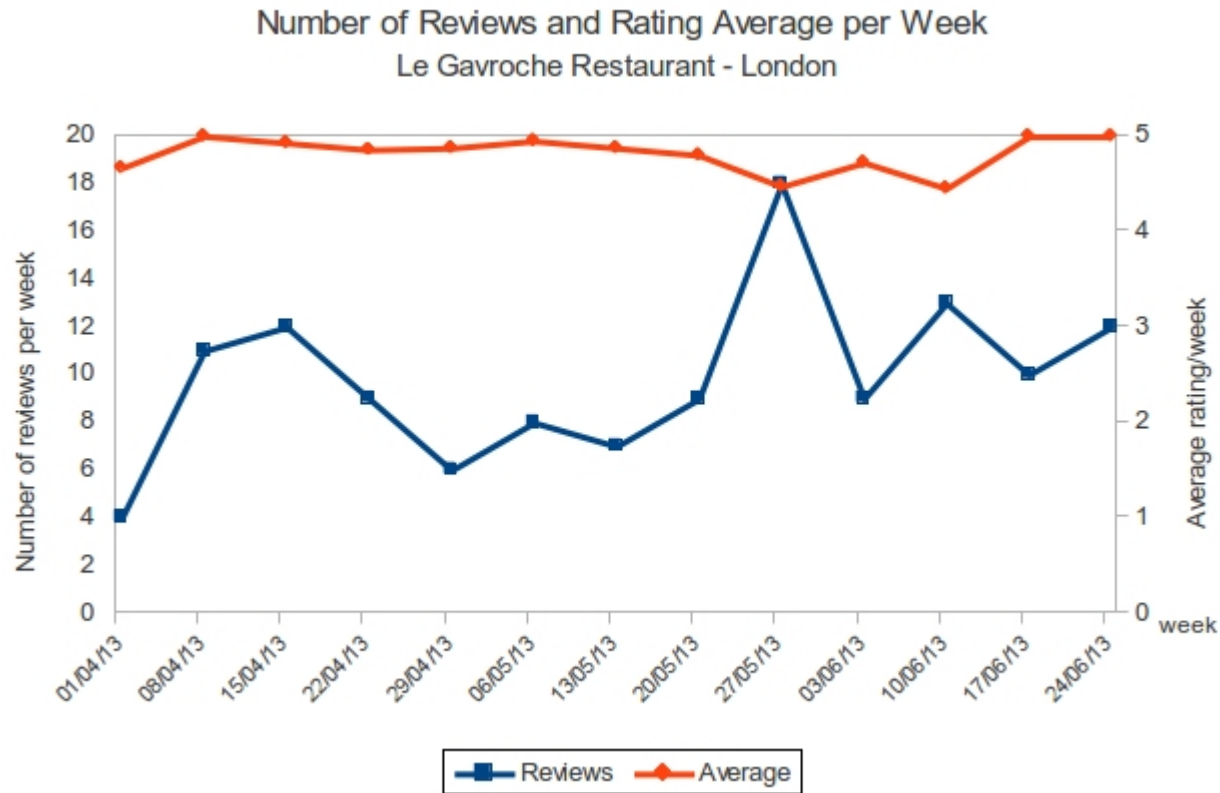
Open Issues

- From a Web Systems point of view, the two main issues around user-generated online content are:
  - The credibility of the published resources
  - The sustainability

Perfect experiences in TripAdvisor

Number of Reviews and Rating Average per Week
Le Gavroche Restaurant - London

Conclusions

- Tim Berners Lee in "Weaving the Web" envisaged the web as both readable and writable, and this has definitely come to pass in the last decade.

- Sustaining the future evolution of web systems will be as much a technological and engineering problem as a social one

- We need to keep researching on understanding of how open on-line collaborative communities develop and function effectively to allow them to keep playing a role in driving future web evolution.